

CLAIMS

What is claimed is:

1. A method of automated data mining using a domain-specific analytic application for solving predefined problems, the method comprising:
  - 5 populating input data schema, the input data schema having a format appropriate to solution of a predefined problem;
  - production training a predefined data mining model to produce a trained data mining model, the predefined data mining model comprising a predefined data mining model definition;
  - 10 production scoring input data from the input data schema; and
  - scheduling the steps of populating input data schema, production training, and production scoring.
- 15 2. The method of claim 1 wherein populating input data schema further comprises reading input data from a data store and writing the input data to the input data schema.
3. The method of claim 2 wherein the data store comprises historical data and populating input data schema further comprises populating input data schema with historical data.

4. The method of claim 2 wherein the data store comprises production data and populating input data schema further comprises populating input data schema with production data.

5. The method of claim 1 wherein:

production training has an input comprising the input data stored in the input data schema and an output comprising a knowledge base,

5

production training further comprises executing a preselected data mining algorithm in production training mode,

executing the data mining algorithm in production training mode comprises  
10 executing a software process within the analytic application, and

the trained data mining model comprises the predefined data mining model definition and the knowledge base.

6. The method of claim 1 wherein production scoring further comprises:

applying the trained data mining model by executing the data mining algorithm in production scoring mode, wherein:

5

the data mining algorithm executed in production scoring mode comprises a software process within the analytic application, and



10        executing the data mining algorithm has an output comprising production scored data.

7.        The method of claim 6 further comprising storing the production scored data in a store of production data.

8.        The method of claim 1 wherein scheduling the steps of populating input data schema, production training, and production scoring further comprises:

5        storing a schedule in computer memory; and

      executing the steps of populating input data schema, production training, and production scoring in dependence upon the schedule.

9.        The method of claim 8 wherein storing a schedule further comprises storing the schedule in a cron table for a scheduler comprising a Unix daemon called "cron."

10.       The method of claim 8 wherein storing a schedule further comprises storing the schedule in an IBM DB2 Warehouse Center scheduling table for a scheduler comprising an IBM DB2 Warehouse Center scheduling facility.

11.       The method of claim 1 wherein the analytic application comprises:

5        the predefined problems to be solved, the predefined problems having referents defined in a specific computational domain;

      predefined data mining algorithms capable of using input data read from predefined

input data schema for solving the predefined problems;

10 predefined data schema appropriate for solution of the predefined problems, the  
predefined data schema further comprising the input data schema and output data  
schema; and

at least one predefined data mining model definition, the predefined data mining  
model definition dependent upon the predefined data schema.

15

12. The method of claim 11 wherein the predefined data mining algorithms comprise a  
radial basis function algorithm for value prediction and the production scored data  
comprises a prediction field containing a value calculated by executing the data  
mining algorithm in scoring mode.

5

13. The method of claim 11 wherein the predefined data mining algorithms comprise a  
neural value prediction algorithm.

14. The method of claim 11 wherein the predefined data mining algorithms comprise a  
demographic clustering algorithm.

15. The method of claim 11 wherein the predefined data mining algorithms comprise a  
neural clustering algorithm.

16. The method of claim 11 wherein the predefined data mining algorithms comprise a  
tree classification algorithm.

17. The method of claim 11 wherein the predefined data mining algorithms comprise a

neural classification algorithm.

18. The method of claim 11 wherein the predefined data mining algorithms comprise an associations algorithm.
19. The method of claim 1 wherein the domain-specific application operates under a first computer operating system and populating input data schema further comprises executing a software program that operates under the first computer operating system.
- 5 20. The method of claim 1 wherein the domain-specific application operates under a first computer operating system and populating input data schema further comprises executing a software program that operates under a second computer operating system.
- 5 21. The method of claim 1 wherein the domain-specific application comprises a software program capable of populating input data schema and populating input data schema further comprises executing from within the domain-specific analytic application the software program capable of populating input data schema.
- 5 22. The method of claim 1 wherein the input data from the input data schema in production scoring further comprises production data.
23. The method of claim 1 wherein the input data from the input data schema in production scoring further comprises historical data.

24. A system for automated data mining using a domain-specific analytic application for solving predefined problems, the system comprising:

5 means for populating input data schema, the input data schema having a format appropriate to solution of a predefined problem;

means for production training a predefined data mining model to produce a trained data mining model, the predefined data mining model comprising a predefined data mining model definition;

10 means for production scoring input data from the input data schema; and

means for scheduling populating input data schema, production training, and production scoring.

- 15 25. The system of claim 24 wherein means for populating input data schema further comprises means for reading input data from a data store and means for writing the input data to the input data schema.
26. The system of claim 25 wherein the data store comprises historical data and means for populating input data schema further comprises means for populating input data schema with historical data.
27. The system of claim 25 wherein the data store comprises production data and means for populating input data schema further comprises means for populating input data schema with production data.

28. The system of claim 24 wherein:

means for production training has an input comprising the input data stored in the input data schema and an output comprising a knowledge base,

5

means for production training further comprises means for executing a preselected data mining algorithm in production training mode,

means for executing the data mining algorithm in production training mode

10

comprises means for executing a software process within the analytic application, and

the trained data mining model comprises the predefined data mining model definition and the knowledge base.

15

29. The system of claim 24 wherein means for production scoring further comprises:

means for applying the trained data mining model by executing the data mining algorithm in production scoring mode, wherein:

5

the data mining algorithm executed in production scoring mode comprises a software process within the analytic application, and

means for executing the data mining algorithm has an output comprising production scored data.

10

30. The system of claim 29 further comprising means for storing the production scored data in a store of production data.

31. The system of claim 24 wherein means for scheduling populating input data schema, production training, and production scoring further comprises:

means for storing a schedule in computer memory; and

5

means for executing the steps of populating input data schema, production training, and production scoring in dependence upon the schedule.

32. The system of claim 31 wherein means for storing a schedule further comprises means for storing the schedule in a cron table for a scheduler comprising a Unix daemon called "cron."

33. The system of claim 31 wherein means for storing a schedule further comprises means for storing the schedule in an IBM DB2 Warehouse Center scheduling table for a scheduler comprising an IBM DB2 Warehouse Center scheduling facility.

34. The system of claim 24 wherein the analytic application comprises:

the predefined problems to be solved, the predefined problems having referents defined in a specific computational domain;

5

predefined data mining algorithms capable of using input data read from predefined input data schema for solving the predefined problems;



10 predefined data schema appropriate for solution of the predefined problems, the predefined data schema further comprising the input data schema and output data schema; and

at least one predefined data mining model definition, the predefined data mining model definition dependent upon the predefined data schema.

15

35. The system of claim 33 wherein the predefined data mining algorithms comprise a radial basis function algorithm for value prediction and the production scored data comprises a prediction field containing a value calculated by means for executing the data mining algorithm in scoring mode.

5

36. The system of claim 33 wherein the predefined data mining algorithms comprise a neural value prediction algorithm.

37. The system of claim 33 wherein the predefined data mining algorithms comprise a demographic clustering algorithm.

38. The system of claim 33 wherein the predefined data mining algorithms comprise a neural clustering algorithm.

39. The system of claim 33 wherein the predefined data mining algorithms comprise a tree classification algorithm.

40. The system of claim 33 wherein the predefined data mining algorithms comprise a neural classification algorithm.

41. The system of claim 33 wherein the predefined data mining algorithms comprise an associations algorithm.

42. The system of claim 24 wherein the domain-specific application operates under a first computer operating system and means for populating input data schema further comprises means for executing a software program that operates under the first computer operating system.

5

43. The system of claim 24 wherein the domain-specific application operates under a first computer operating system and means for populating input data schema further comprises means for executing a software program that operates under a second computer operating system.

5

44. The system of claim 24 wherein the domain-specific application comprises a software program capable of populating input data schema and means for populating input data schema further comprises means for executing from within the domain-specific analytic application the software program capable of populating input data schema.

5

45. The system of claim 24 wherein the input data from the input data schema in production scoring further comprises production data.

46. The system of claim 24 wherein the input data from the input data schema in production scoring further comprises historical data.

47. A computer program product for automated data mining using a domain-specific analytic application for solving predefined problems, the computer program product comprising:

5

a recording medium;

means, recorded on the recording medium, for populating input data schema, the input data schema having a format appropriate to solution of a predefined problem;

10

means, recorded on the recording medium, for production training a predefined data mining model to produce a trained data mining model, the predefined data mining model comprising a predefined data mining model definition;

15

means, recorded on the recording medium, for production scoring input data from the input data schema; and

means, recorded on the recording medium, for scheduling populating input data schema, production training, and production scoring.

20

48. The computer program product of claim 47 wherein means for populating input data schema further comprises means, recorded on the recording medium, for reading input data from a data store and means, recorded on the recording medium, for writing the input data to the input data schema.

5

49. The computer program product of claim 48 wherein the data store comprises historical data and means for populating input data schema further comprises

means, recorded on the recording medium, for populating input data schema with historical data.

5

50. The computer program product of claim 48 wherein the data store comprises production data and means for populating input data schema further comprises means, recorded on the recording medium, for populating input data schema with production data.

5

51. The computer program product of claim 47 wherein:

means for production training has an input comprising the input data stored in the input data schema and an output comprising a knowledge base,

5

means for production training further comprises means, recorded on the recording medium, for executing a preselected data mining algorithm in production training mode,

10

means for executing the data mining algorithm in production training mode comprises means, recorded on the recording medium, for executing a software process within the analytic application, and

the trained data mining model comprises the predefined data mining model definition and the knowledge base.

15

52. The computer program product of claim 47 wherein means for production scoring further comprises:

means, recorded on the recording medium, for applying the trained data mining  
5 model by executing the data mining algorithm in production scoring mode,  
wherein:

the data mining algorithm executed in production scoring mode comprises a  
software process within the analytic application, and  
10

means for executing the data mining algorithm has an output comprising production  
scored data.

53. The computer program product of claim 52 further comprising means, recorded on  
the recording medium, for storing the production scored data in a store of  
production data.

54. The computer program product of claim 47 wherein means for scheduling  
populating input data schema, production training, and production scoring further  
comprises:

5 means, recorded on the recording medium, for storing a schedule in computer  
memory; and

means, recorded on the recording medium, for executing the steps of populating  
input data schema, production training, and production scoring in dependence upon  
10 the schedule.

55. The computer program product of claim 54 wherein means for storing a schedule  
further comprises means, recorded on the recording medium, for storing the

schedule in a cron table for a scheduler comprising a Unix daemon called "cron."

56. The computer program product of claim 54 wherein means for storing a schedule further comprises means, recorded on the recording medium, for storing the schedule in an IBM DB2 Warehouse Center scheduling table for a scheduler comprising an IBM DB2 Warehouse Center scheduling facility.

5

57. The computer program product of claim 47 wherein the analytic application comprises:

the predefined problems to be solved, the predefined problems having referents defined in a specific computational domain;

5

predefined data mining algorithms capable of using input data read from predefined input data schema for solving the predefined problems;

10

predefined data schema appropriate for solution of the predefined problems, the predefined data schema further comprising the input data schema and output data schema; and

at least one predefined data mining model definition, the predefined data mining model definition dependent upon the predefined data schema.

15

58. The computer program product of claim 57 wherein the predefined data mining algorithms comprise a radial basis function algorithm for value prediction and the production scored data comprises a prediction field containing a value calculated by means, recorded on the recording medium, for executing the data mining algorithm

5 in scoring mode.

59. The computer program product of claim 57 wherein the predefined data mining algorithms comprise a neural value prediction algorithm.
60. The computer program product of claim 57 wherein the predefined data mining algorithms comprise a demographic clustering algorithm.
61. The computer program product of claim 57 wherein the predefined data mining algorithms comprise a neural clustering algorithm.
62. The computer program product of claim 57 wherein the predefined data mining algorithms comprise a tree classification algorithm.
63. The computer program product of claim 57 wherein the predefined data mining algorithms comprise a neural classification algorithm.
64. The computer program product of claim 57 wherein the predefined data mining algorithms comprise an associations algorithm.
65. The computer program product of claim 47 wherein the domain-specific application operates under a first computer operating system and means for populating input data schema further comprises means, recorded on the recording medium, for executing a software program that operates under the first computer operating  
5 system.
66. The computer program product of claim 47 wherein the domain-specific application

operates under a first computer operating system and means for populating input data schema further comprises means, recorded on the recording medium, for executing a software program that operates under a second computer operating system.

5

67. The computer program product of claim 47 wherein the domain-specific application comprises a software program capable of populating input data schema and means for populating input data schema further comprises means, recorded on the recording medium, for executing from within the domain-specific analytic application the software program capable of populating input data schema.

5

68. The computer program product of claim 47 wherein the input data from the input data schema in production scoring further comprises production data.

69. The computer program product of claim 47 wherein the input data from the input data schema in production scoring further comprises historical data.

5